PREEMPTION AND A DILEMMA FOR CAUSAL DECISION THEORY

Esteban Céspedes
Goethe University, Frankfurt am Main, Germany
E-mail: e.cespedes@stud.uni-frankfurt.de

Abstract

One of the lessons given by the prisoners' dilemma is that collective decisions are more rational when they are based not only on evidence, but also on causal relations. This is solved by causal decision theory. However, the notion of causation this theory is based on confronts further problems in preemption cases. It will be shown briefly that preemption does not occur less frequently in social and economic situations than in prisoners' dilemma and usual causal scenarios. Group decision theory and competition are clear (and perhaps not the only) examples of that. It will be argued that, in order to solve a so called *preemption dilemma*, the smallest theoretical alteration should be focused on preemption rather than on the dilemma. Amongst the most relevant approaches, structural equation models and ranking analysis of causation provide appropriate answers.

Keywords: decision, causation, prisoner's dilemma, preemption, counterfactual.

1. The prisoner's dilemma

The dilemma has taken many shapes since its original formulations [1]. New formulations [2] will be better suited for present purposes, which are based only on the one-attempt version of the problem, letting aside its iterated versions. Two prisoners situated in unconnected cells have the same two options: stay silent or talk about their crime. Further, they are aware of the following conditions: if both decide to talk, then both will get ten years of prison; if both decide to stay silent, both get one year of prison; if one of them talks and the other stays silent, then the first one goes free and the other gets a sentence of twenty years. A matrix of the situation has this form:

$$\begin{array}{lll} & \mathbf{S} & \mathbf{T} \\ \mathbf{S} & 1 & 20 \\ \mathbf{T} & 0 & 10 \end{array} \qquad (1)$$

Here the rows represent the options: staying silent (S) or talk (T). The columns represent the possible outcomes given the other prisoner's decision. A good manner to analyse this problem is adopting, as agent, one of the two prisoners' perspective. The figure shows that the best possible outcome for the agent can only occur if he stays silent. Thus, the dominant option is talking rather than cooperating. Nevertheless, once the agent thinks that the other prisoner might come to the same conclusion, the outcome seems to be not so good.

2. Causation and the prisoners

This dilemma can take many forms. It is well known, e.g., that prisoners' dilemma is closely connected to Newcomb's problem. Some authors [3] have shown that both are the same problem, while others [4] have argued that not every formulation of Newcomb's problem is a prisoners' dilemma. Since Newcomb's problem is better understood under a causal notion of decision [5][6], one might need to introduce causal relations into some versions of the prisoners' dilemma.

One important element of introducing causation into decision analysis is connected with the distinction between *evidential* and *causal* decision theories [7]. While the former suggest that an agent's decision can be considered as part of the evidence for certain outcome, the latter are based on the causal dependence between the agent and the outcome. According to evidential decision theory, the agent should maximise the expected value of his options, conditionalising the probability of the possible outcomes (O) on his action (A):

$$V(\mathrm{A}) = \Sigma_j P(\mathrm{O_j}|\mathrm{A})V(\mathrm{O_j}) \qquad (2)$$

The maximisation of expected value thus defined leads to irrational decisions, like preferring to avoid the medical examination because it will lower the probability of a bad diagnose. The criterion used in definition (2) suggests the agent to change his mind, because his decision might be an evidence to think that the other prisoner is going to do the same. Causal decision theory replaces conditional probabilities with counterfactual conditionals, in order to define *expected utility*, which should be maximised in rational decisions. The difference between the words 'value' and 'utility' is based on at least three reasons. The first is the continuation of the original notation used by Gibbard and Harper [7], which actually differentiates between two kinds of utility, $\mathcal{V}$-utility and $\mathcal{U}$-utility, instead of considering the notion of *value*. A second reason comes from the ordinary meaning of 'utility', which is near to the conceptions of causal *usefulness* [5] and *efficacy* [7]. The third reason has to do with the more quantitative character of $\mathcal{V}$-utility that depends on conditional probability while $\mathcal{U}$-utility is defined with a counterfactual conditional. Let '$\Box\!\!\rightarrow$' be the counterfactual operator:

$$\mathrm{U(A)} = \Sigma_{\mathrm{j}}\, P(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{O_j})\, V(\mathrm{O_j}) \qquad (3)$$

Adding a total, or in some cases just practical, pattern of causal dependence would define expected utility in this way [5]:

$$\mathrm{U(A)} = \Sigma_{\mathrm{K}}\, P(\mathrm{K})\, V(\mathrm{AO}) \qquad (4)$$

The dependency hypothesis (K) describes how things are causally influenced by the agent's actions using counterfactual sentences. These sentences express causal dependence according to the counterfactual definition of causation [8]. Let $\rightsquigarrow$ be the binary relation of causal dependence:

$$\mathrm{C} \rightsquigarrow \mathrm{E} \leftrightarrow \quad \begin{array}{l} \text{a) } \mathrm{C} \mathrel{\Box\!\!\rightarrow} \mathrm{E} \text{ and} \\ \text{b) } \sim\!\mathrm{C} \mathrel{\Box\!\!\rightarrow} \sim\!\mathrm{E} \end{array} \qquad (5)$$

Counterfactual dependence is not transitive, whence causal dependence cannot be transitive either. However, the transitive relation of causation can be constructed through a chain of causal dependencies [8]. Dependency hypotheses permit the agent to ignore the occurrences that escape his control in order to achieve his goals. That suggests him to confess in prisoners' dilemma, because what the other prisoner decides does not depend causally on what *he* decides, i.e. it is not a consequence of the dependence hypothesis. In the following section I will briefly introduce the structural model of causation, which does not replace definition (5), but reinforces it. Although the variation between different versions of causal decision theory is not huge, it has been shown that considering these details might be of great relevance [9].

3. Structural models of decision

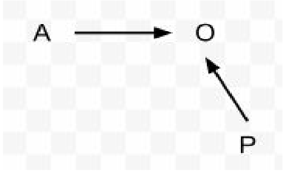
Dependency hypotheses take the form of a set of structural equations or *parameters* over the variables under consideration in structural accounts for causality [6]. This causal structure of the set of variables is represented as a directed acyclic graph, which relates the variables as nodes through causal dependence. In this framework the causal model is formed by a set of functional equations, which describes the dependencies of the variables on their so called *parents* and disturbances or unobserved variables of the system. Through the notion of *intervention*, the model can be modified counterfactually without changing the values of the other variables. In this sense, the agent's *action* cannot serve as evidence for a certain outcome in his decision nor can it be taken as one more variable of the model. This approach defines expected utility using interventions (*do*) on a causal model, which represent the action of the agent:

$$\mathrm{U(A)} = \Sigma_{\mathrm{j}}\, P(\mathrm{O_j}|do(\mathrm{A}))\, V(\mathrm{O_j}) \qquad (6)$$

The definition for expected utility considers the potential influence that an action might have on the causal model expressed in counterfactual conditions. Although the outcome is taken to be conditioned in a probabilistic way by the intervention, it does not mean that the definition is similar to the expected value explained in the equation (2). The *do* operator has already a counterfactual sense. Actually these are equivalent notions [6]:

$$P(\mathrm{y}|do(\mathrm{x})) = P[(\mathrm{X{=}x}) \mathrel{\Box\!\!\rightarrow} (\mathrm{Y{=}y})] \qquad (7)$$

Definition (7) says that the probability that *y* occurs given the intervention *x* is equal to the probability of a counterfactual conditional: if the variable *X* had the value *x*, the variable *Y* would have the value *y*. In order to reveal the causal dependences and independences of the prisoners' dilemma, a causal graph should be constructed connecting the variables in play. Let *A* be the two-valued action variable that stands for the agent's possible options, *O* be the four-valued occurrence variable standing for the possible outcomes, which also depend on *P*, the two-valued occurrence variable standing for the other prisoner and his options.



(8)

Now let *t* be the value 'the agent talks' and *s* be the value 'the agent stays silent'. The outcome has one value for each sentence expressed in number of years. The set of equations that defines this model is the following:

$$\begin{array}{ll} A = & t \vee s \\ P = & t \vee s \\ O = & 0 \vee 1 \vee 10 \vee 20 \\ & \\ O = & 0 \textit{ if } (A=t \wedge P=s) \\ & 1 \textit{ if } (A=s \wedge P=s) \\ & 10 \textit{ if } (A=t \wedge P=t) \\ & 20 \textit{ if } (A=s \wedge P=t) \end{array} \qquad (9)$$

Since A and P are independent variables and the outcome depends causally not only on what the agent does, but also on what the prisoner decides, the most rational decision according to definition (6) would be to talk. The agent's decision only has a causal influence on the outcome, not on the other prisoner. The latter does not causally influence the agent's decision either, which means that, under causal decision theory, the agent should neither consider the other prisoner's possible act nor his decision itself as evidence in favour of a certain outcome. The agent is absolutely uncertain about what the other prisoner is going to choose.

However, the alternative of staying silent in the dilemma comes from the thought that there is somehow a correlation between what the agent does and what the other prisoner does, just because each one considers that the other thinks in the same way he does. But that relation would not be causal; *ex hypothesi* both prisoners are in different cells that are unconnected to each other, which means also *causally* unconnected. Hence, decision theories based on causal dependence dismiss such alternative.

4. Ranking and decisions

Another account, very similar to the described above, defines causal relations through subjective ranking functions [10]. These are functions of disbelief, which go from the set of situations or outcomes to the set of non-negative integers. Thus, giving high ranking to an outcome means that it is highly disbelieved to be the case. The higher the ranking of the outcome, the less likely it would be for an agent to occur. Notice the issue that arises in this account of subjective probabilities after conditionalising; conditional probabilities have trouble when extreme low values are considered. That won't be a problem for the analysis of decision, though, since it would be strange (or useless) to give possible actions an extreme low probability. It is argued that the structural models for causation can be transferred to the causal account of ranking functions, since both are based on the same laws for probabilistic independence [10]. The manner in which both approaches are related is more than interesting, but further similarities will not be analysed here.

Sufficient and necessary causation can be defined using ranking functions. Let ϰ be the ranking function and B be the proposition about some set of obtaining circumstances:

$$C \rightsquigarrow E \leftrightarrow \varkappa(E|C\cap B) < \varkappa(E|{\sim}C\cap B) \qquad (10)$$

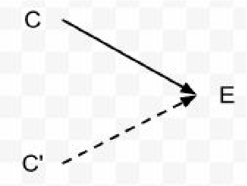
The difference between sufficient and necessary causes lies in how high is the value given to a certain ranking, e.g. if the quantity of the right side of definition (10) is near to the one of the left side, the proposition in question might express a sufficient cause for its effect, while it might be a necessary cause, if the difference between both sides is too high.

This account could show that the optimal decision in prisoners' dilemma is to stay silent and cooperate, if the situation is repeated many times [11]. Under those presuppositions it might be reasonable to think that somehow the agent's and the other prisoner's actions influence each other, but such kind of versions are outside the scope of this work. In any case, it would not be singular causation what connects them. It has been shown elsewhere that earlier accounts of counterfactual causation, like the one explained in definition (5) are special cases of the ranking account for causation [10][12]. Unfortunately, this reduction can only succeed if a further condition is introduced that imposes the precedence of the cause over its effect. This supposition is neither needed in original counterfactual causation nor in the structural model account. Anyway, the relevance of taking counterfactual causation into ranking causation lies on the solution the latter offers to the problem I will describe in the next section.

Another benefit of this account is its subjective character, which adjusts better to the proposals of decision analysis. The definition (10) could serve perfectly to construct a dependency hypothesis in the sense of definition (4) in order to provide a notion of expected utility.

5. Preemption and many agents

The approaches to decision theory that have been mentioned until now are not only causal, but also very much connected to counterfactuals. Every causal decision theory based on a counterfactual account of causation might suffer from the same difficulties that the latter has been confronted with. Preemption is one of them. This problem—discussed very often nowadays [13]—arises when, among two potential causes of an event, the actual one succeeds because it interrupts the other. A causal graph might represent this well. Let *C* be the actual cause of *E*, and *C'* be the potential cause, i.e. an event that could have caused *E*, if *C* had not occurred:



(11)

The definition (5), shown above, does not detect the actual cause in such situations, because the second counterfactual conditional does not obtain; if the actual cause had not been the case, then it is not true that the effect would not have been the case, because the back-up cause was there to produce it. The preemption problem is a kind of overdetermination. However, cases of symmetric overdetermination—i.e. where two or more events taken together are the cause and each of them is necessary to produce the effect—might be explained in the same way as preemption [8].

A very usual scenario of preemption happens in group decision theory, when an individual agent comes up with a solution before his partners, but nevertheless the decision is taken by the whole group. Consensus is a model for decision theory that seeks to avoid sharp preemption [14], which might reduce decisional dictatorship. However, when time is a relevant factor, agents might reach the consensus that the first solution that fulfills certain conditions is the one to be applied in the final decision.

Competition is another social example of causal preemption, where two or more agents have access to the same opportunity and the first to take it leaves it unavailable [15]. Most cases of this kind leave the competitors' costs of taking the decision unknown to the agent, which does not occur in the prisoners' dilemma, because the agent knows that the possible sentences are distributed in the same way for him that for the other prisoner.

If dependency hypotheses are constructed on a counterfactual base, which gets difficulties from preemption cases, then there could be situations where the agent tries to decide according to expected utility and still obtains irrational outcomes after all. A decision analysis that does not tackle preemption carefully might end up suggesting that an agent should not work towards an idea, because his reflections are not going to be the cause of the final group decision either way, even if he comes up with the idea first. Other practical absurdities may also follow with respect to competition.

6. Preemption and the prisoners

Preemption cases undoubtedly pose a lot of problems for causal decision theory and they could raise further difficulties because of the ways in which such account handles prisoners' dilemma. Since that is one of the best benefits of causality-based decisions, cases conjoining both problems are more than threatening for such theory.

A modification in the prisoners' dilemma can generate such a situation. Notice that preemption in decision is not necessarely a prisoner's dilemma. In the conjoined case, however, consider the same information given to the agent represented by figure (1) and the set of equations (9), and change one of the conditions, such that if both prisoners talk, the one who confessed first gets a sentence of five years, and the second gets ten. A table of the mixed problem can be shown:

	S	**T**
S	1	20
T	0	5∨10

(12)

The causal parameters described by the structural equations cannot be the same as in the set of equations (9). Let two new conditions be introduced in exchange of the equation that describes the outcome taking the value of ten years of prison:

$$\begin{aligned} A &= t \vee s \\ P &= t \vee s \\ O &= 0 \vee 1 \vee 5 \vee 10 \vee 20 \end{aligned}$$

$$O = \begin{cases} 0 \;\; \textit{if} \;\; (A{=}t \wedge P{=}s) \\ 1 \;\; \textit{if} \;\; (A{=}s \wedge P{=}s) \\ 5 \;\; \textit{if} \;\; (A{=}t_1 \wedge P{=}t_2) \\ 10 \;\; \textit{if} \;\; (A{=}t_2 \wedge P{=}t_1) \\ 20 \;\; \textit{if} \;\; (A{=}s \wedge P{=}t) \end{cases} \qquad (13)$$

In the two new conditions, subscripts stand for the order in which the confessions took place. If the agent confesses first, he gets only five years of prison. If he talks late, he will get the same sentence as in the original dilemma. The causal graph stays the same as in figure (11), except that the variable O becomes five-valued.

In a *preemption dilemma* the agent must act quickly, which means that he should look for preemption. The optimal decision is no longer the one that just expects the other prisoner to stay silent, but the one who seeks to preempt the other prisoner's act. But in order to maximise the utility of one's action in such situations, preemption must be explained with clarity and considering the relevant causal relations. Otherwise, the agent is likely to ignore that he influences the variable of interest with certain detail. Thus, a dependency hypothesis based on naïve causal counterfactuals would fail in front of decision cases where preemption is optimal. It will construct a set of equations like the one described in (9), instead of considering a set like (13), because the agent won't be able to judge that the way or time in which he decides modifies the pattern of causal influences. These considerations give reason to establish that whether the effect occurred does not only depend on whether the cause occurred, but how and when it occurred [16]. Nevertheless, such a solution might lead to indistinguishable spurious causes [8]. It has also been established that a poor understanding of causation in competitive cases might lead to irrelevance in decisions, waste of efforts, or even to epistemological mistakes [17].

Other ways of confronting preemption are proposed by the structural model account and by the ranking analysis of causation. The first approach introduces the notion of sustenance, which emphasises the capacity of the actual cause to maintain the value of the effect when some structural contingences are suppressed [6, p. 316]. Thus, the property of sustenance lies between the concepts of production and dependence. On the other hand, the ranking theory of causation manages the problem of overdetermination in general appealing to the notion of *additional* cause [10, p. 110]. A table of rankings of the outcome conditioned on the different causes shows the difference:

ϰ(O| .)	***C***	***~C***	
C'	*0*	*1*	
~C'	*0*	*2*	*(14)*

The value of the effect's occurrence in absence of an additional cause is not much higher than in its presence. In preemption, the occurrence of the effect gets higher ranks when the actual cause fails to occur than when the backup cause does. Both sustenance and the distinction of additional causes are notions that provide precise responses to preemption in general, and to more particular cases of preemption dilemmas when considered in decision theoretic contexts.

7. Conclusion

Some accounts of decision based on causal relations have been briefly presented, showing that the original approach handles well the prisoners' dilemma [5][7], but that it might lead to irrationality if confronted with preemption cases. This kind of problem is called preemption dilemma and I am sure that it is a real problem for causal decision theory. Solutions were considered and recommended under the scope of structural models and ranking theory of causality. It is important that such solutions maintain what causal decision theory has done for the prisoners' dilemma. In that way the difficulties can be avoided trough the smallest possible change in the theory. After all, the trouble for causal decision theory generated by preemption dilemmas does not lie on the side of the dilemma, but on the side of how preemption should be understood under decision analyses.

References

[1] Poundstone, W. (1992). *Prisoner's Dilemma: John Von Neumann, Game Theory and the Puzzle of the Bomb*. Doubleday, New York.

[2] Schick, F. (2004). A Dilemma for Whom? *Synthese* 140: 3-16.

[3] Lewis, D. (1986). Prisoners' Dilemma is a Newcomb Problem. *Philosophical Papers Volume II*, 1(9):299-305.

[4] Sobel, J. (1991). Some Versions of Newcomb's Problem are Prisoners' Dilemmas. *Synthese,* 86: 197-208.

[5] Lewis, D. (1981). Causal Decision Theory. *Australasian Journal of Philosophy*, 59(1):5-30.

[6] Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York.

[7] Gibbard, A. & Harper, W. (1978). Counterfactuals and Two Kinds of Expected Utility. In Hooker, Leach & McClennen (eds.). *Foundations and Applications of Decision Theory*. Reidel, Dordrecht.

[8] Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17):556-567.

[9] Rabinowicz, W. (1982). Two Causal Decision Theories: Lewis vs. Sobel. In Tom, P. *Philosophical Essays Dedicated to Lennart Åqvist*. Philosophical Studies, Uppsala.

[10] Spohn, W. (2006). Causation: An Alternative. *British Journal for the Philosophy of Science*, 57(1):93-119.

[11] Spohn, W. (2007). Dependency Equilibria. *Philosophy of Science* 74 (5):775-789.

[12] Huber, F. (2011). Lewis Causation is a Special Case of Spohn Causation. *British Journal for the Philosophy of Science,* 62 (1):207-210.

[13] Collins, J. (2004). Preemptive Prevention. In Collins, Hall & Paul (eds.). *Causation and Counterfactuals*. MIT, Massachussetts.

[14] Caws, P. (1991). Committees and Consensus: How Many Heads Are Better Than One? *Journal of Medicine and Philosophy,* 16 (4).

[15] Anderson, S., Friedman, D., and Oprea, R. (2010). Preemption Games: Theory and Experiment. *American Economic Review*, 100(4): 1778-1803.

[16] Lewis, D. (2000). Causation as Influence. *The Journal of Philosophy*, 97(4):182-197.

[17] Durand, R. and Vaara, E. (2009). Causation, Counterfactuals, and Competitive Advantage. *Strategic Management Journal*, 30(12):1245-1264.

Acknowledgements

I am deeply indebted to André Fuhrmann, José Tomás Alvarado, Rafael Miranda and Manfred Kupffer for iluminating comments and criticism and to the Konrad Adenauer Foundation for general support to this research.